

DOI: 10.1515/jwld-2016-0021

© Polish Academy of Sciences, Committee for Land Reclamation
and Environmental Engineering in Agriculture, 2016
© Institute of Technology and Life Sciences, 2016

JOURNAL OF WATER AND LAND DEVELOPMENT
2016, No. 30 (VII–IX): 51–56
PL ISSN 1429–7426

Available (PDF): <http://www.itp.edu.pl/wydawnictwo/journal>; <http://www.degruyter.com/view/j/jwld>

Received 28.04.2016
Reviewed 25.05.2016
Accepted 08.06.2016

A – study design
B – data collection
C – statistical analysis
D – data interpretation
E – manuscript preparation
F – literature search

Time series analysis of reference crop evapotranspiration for Bokaro District, Jharkhand, India

Ratnesh GAUTAM^{ABC}, Anand K. SINHA^{DEF}

Birla Institute of Technology, Department of Civil and Environmental Engineering, Mesra-835215, India;
e-mail: ratnesh_mishraji@yahoo.co.in, aksinha@bitmesra.ac.in

For citation: Gautam R., Sinha A.K. 2016. Time series analysis of reference crop evapotranspiration for Bokaro District, Jharkhand, India. *Journal of Water and Land Development*. No. 30 p. 51–56. DOI: 10.1515/jwld-2016-0021.

Abstract

Evapotranspiration is the one of the major role playing element in water cycle. More accurate measurement and forecasting of Evapotranspiration would enable more efficient water resources management. This study, is therefore, particularly focused on evapotranspiration modelling and forecasting, since forecasting would provide better information for optimal water resources management. There are numerous techniques of evapotranspiration forecasting that include autoregressive (AR) and moving average (MA), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), Thomas Feiring, etc. Out of these models ARIMA model has been found to be more suitable for analysis and forecasting of hydrological events. Therefore, in this study ARIMA models have been used for forecasting of mean monthly reference crop evapotranspiration by stochastic analysis. The data series of 102 years i.e. 1224 months of Bokaro District were used for analysis and forecasting. Different order of ARIMA model was selected on the basis of autocorrelation function (ACF) and partial autocorrelation (PACF) of data series. Maximum likelihood method was used for determining the parameters of the models. To see the statistical parameter of model, best fitted model is ARIMA (0, 1, 4) (0, 1, 1)₁₂.

Key words: ARIMA model, evapotranspiration, forecasting, time series

INTRODUCTION

Water is the precious gift of nature which requires to sustaining life on earth. To see the increasing demand of population and climate change, water need to be managing carefully and precisely. For planning, managing, and operation of water resources projects, irrigation scheduling, to make local and national water policy etc., evapotranspiration estimation is important. Evapotranspiration rate from a reference surface, not short of water, is known as reference crop evapotranspiration and is denoted by ET_0 (FAO 56). For analysis of evapotranspiration numbers of exiting methods are available. In this study time series analysis of evapotranspiration is to be done. A data set of equal time interval is known as time series. A time

series is generally divided in two components i.e. deterministic and stochastic. Deterministic component is applied for forecasting of time and chance independent future events, whereas stochastic component is used for determination of chance and chance dependent effects. Deterministic component shows increasing or decreasing trends in data series. The periodic nature of deterministic component is classified by its cyclic pattern, which is repeated over fix interval of time. The stochastic component contains irregular oscillation and random effect, and is defined by probabilistic concept. To avoid erratic result, a time series must exhibit smooth behaviour. The most popular time series model is the autoregressive integrated moving average (ARIMA) model. In this model the forecast of a variable is defined as a linear combina-

tion of the previous state of variable (pure AR component), and previous forecast error (pure MA component). Integrated means differencing the data series for removing seasonal or periodic component. ALHASSOUNL *et al.* [1997] developed autoregressive model for ten different station of Saudi Arabia. VALIPOUR [2012] analyse the seasonal autoregressive integrated moving average model for Mehrabad Synoptic Station, Tehran, Iran. MOHAN and ARUMUGAM [1995] studied on seasonal ARIMA modelling of weekly data of evapotranspiration of Annamainagar Meteorological Station, India. VALIPOUR [2012] used ARIMA model for forecasting for Mehrabad Synoptic Station, Teharan, Iran. PARERA *et al.* [2013] analysed and forecast of daily *ET_o*. POPALE and GORANTIWAR [2014] used ARIMA model for forecasting rainfall of Rahuri Region, India. GORANTIWAR and PATIL [2009] did analysis of *ET_o* of Rahuri Region, India. XU *et al.* [2012] predicted *ET_o* of Nanjing Climate Station, China. PSILOVIKOS and ELHAG [2013] forecasted remotely sensed daily *ET_o* of Nile Delta Region, Egypt. KOTTEGODA [1980] explained ARIMA and seasonal ARIMA very clearly. ETUK and NATAMBA [2015] modelled seasonal Box-Jenkins for monthly Ugandan shilling/US Dollar. HAMDY *et al.* [2008] develop seasonal ARIMA Model for the Jordan valley. ASADI *et al.* [2014] forecasts evapotranspiration for humid and semi-humid region. SALAS *et al.* [1980] discussed in detail about time series modelling. The objective of this study is to establish a time series model to analyse and forecast reference crop evapotranspiration for the Bokaro District.

MATERIAL AND METHODS

Mean monthly data of reference crop evapotranspiration (*ET_o*) for the span of 102 years (1901–2002) are collected from India Waterportal [dateless]. Data is of Bokaro District, Jharkhand State, India in $\text{mm}\cdot\text{day}^{-1}$. Bokaro is situated at latitude 23.6, longitude 86.1, and having area approximately 2861 km^2 . Surrounding districts are Ramgarh District to west, Dhanbad District to the east, Puruliya District to the south. It is located 112 km state capital Ranchi. Population of Bokaro District is 2 061 918. Average altitude is 210 m m.s.l. Bokaro is too hot in summer, day time temperature in summer is 31°C to 45°C, January average temperature is 18°C, February average temperature is 22°C, March average temperature is 28°C, April is 33°C, May is 36°C. Bokaro has network of valleys and sub-valleys. Geography includes the the Damodar River and its tributaries. They fulfil important source of water for the city. Bokaro Disrcit has large steel plants, so that it is famous by Bokaro Steel City.

A set of equally-spaced time periods $y_1, y_2, y_3 \dots y_n$ is known as time series. Analysis of time series starts with plots of equally spaced points in the time series. Then to know correlation between each point of series to its earlier value autocorrelation and partial autocor-

relation are to be determined. In this study seasonal ARIMA modelling is applied for analysis of reference crop evapotranspiration. ARIMA model consists of three main components, identification of model, parameter estimation, and diagnostic checking [BABAZADEH, SHAMSNLA 2014]. Data set should be stationary before analysis of ARIMA modelling. A time series which mean and the autocorrelation structure are constant over time is known as stationary time series. When a time series contains trend, periodicity, and heteroscedasticity, then differencing and power transformation are generally applied to the data set to remove the trend, periodicity before an ARIMA model can be fitted. Auto correlation and partial autocorrelation determination are required, to know AR and MA present in the model. Parameter estimation consists of to use the data set to estimate parameters of tentatively identified model. At the time of estimating parameter, overall residuals measure should be minimized. This is to be done with a nonlinear optimization technique. Last stage of model building is diagnostic checking. This stage show weather residuals are normally distributed independent, and homoscedastic. Diagnostic checking can help to suggest alternative models. Final selected model can be used for forecasting purpose. Autoregressive Integrated Moving Average (ARIMA) modelling is based on past data behaviour over time. In analysis, observation which is taken at constant interval of time, considered random variables. ARIMA model expose hidden patterns in data and on the basis of past forecasts future. Forecast can become more accurate if seasonal pattern is also eliminated. ARIMA manly contains two parts one is autoregressive (AR), another is moving average (MA). The relationship between present and past observation is shown by AR parts, whereas MA show autocorrelation of error. ARIMA model is define by (p, d, q) $(P, D, Q)_s$. The equations of ARIMA model is following:

$$\Phi_p(B)\phi_p(B^s)(1-B)^d(1-B)^D Z_t = C + \theta_q(B)\Theta_Q(B^s)\epsilon_t, t = 1, 2, n \quad (1)$$

$$\Phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 \dots - \phi_p B^p) \quad (2)$$

$$\Phi_p(B^s) = (1 - \phi_{1,s} B^s - \phi_{2,s} B^{2s} \dots - \phi_{p,s} B^{ps}) \quad (3)$$

$$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 \dots - \theta_q B^q) \quad (4)$$

$$\Theta_Q(B^s) = (1 - \theta_{1,s} B^s - \theta_{2,s} B^{2s} \dots - \theta_{q,s} B^{Qs}) \quad (5)$$

Where Φ is the unknown parameters of autoregressive and θ is the unknown parameter of moving average at different lag of time t . General form of Monthly Seasonal ARIMA model is $(p, d, q)(P, D, Q)_{12}$. In this model p represents the numbers of non-seasonal autoregressive value, d represents order of non-seasonal differencing, q represents numbers of non-seasonal moving average value, P represents numbers of seasonal autoregressive value, D represents order of sea-

sonal differencing, Q represents numbers of moving average value, and s is the order of the seasonal pattern appearing in the data set. In case of monthly data s will become 12. (p, d, q) is called non-seasonal part of model whereas (P, D, Q) is called seasonal part of the model. Primary value of $(p, d, q)(P, D, Q)_{12}$ are determined using ACF and PACF.

Stationary R^2 compares simple mean model to stationary part of model. This measure is usually preferable instead of ordinary R^2 when seasonal pattern appear in data set. The range of stationary R^2 is negative infinity to 1. R^2 also known as coefficient of determination is the proportion of variation in dependent variable explained by regression model. Value of R^2 varies from 0 to 1. Root mean square error ($RMSE$) is the square root of mean square error, which shows how much series varies from its forecasted value. Mean absolute percentage error ($MAPE$), maximum absolute percentage error ($MaxAPE$), mean absolute error (MAE), maximum absolute error ($MaxAE$), normalized Bayesian information criterion (BIC), the Ljung-Box Q statistics are also used for error estimation.

DATA COLLECTION AND ITS ANALYSIS

Data series of 1224 months are plotted on simple graph paper. For removing of periodicity data series was subjected to the seasonal and non-seasonal differencing. The data series is subjected to plot of autocorrelation function (ACF) and partial autocorrelation function (PACF) which is presented in Figure 1 and Figure 2 respectively. The continuous line in the plot shows confidence limit of 95% significance level. ACF value is not significantly vary from zero. Visual inspection clearly represents strong periodicity and some correlation of lag 12. In order to remove the

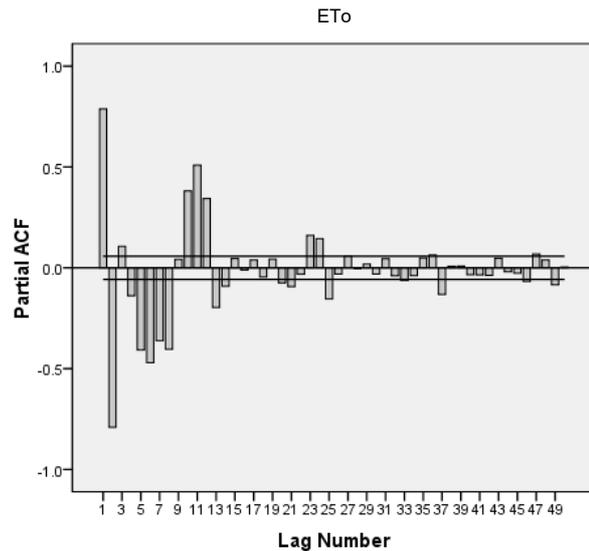


Fig. 2. Partial auto correlation function (PACF) of Bokaro reference crop evapotranspiration (ET_o) up to 50 lags; source: own study

periodicity to make series stationary seasonal and non-seasonal differencing of data series is done. Then ACF and PACF of seasonal differenced data were plotted in Figure 3 and 4. The ACF indicates significant difference zero at first and twelfth lag, showing presence of MA(1), and seasonal MA(1) component of the series. The PACF pattern is also indicating AR(1), and seasonal AR(1). Figure 5 and 6 show ACF&PACF of seasonal and non-seasonal differenced data. They also indicate presence of AR(1) and MA(1), MA(2) seasonal and non-seasonal. To see the different spikes appear in ACF and PACF, different sets of ARIMA models are to be tried in the form of $(p, d, q)(P, D, Q)_s$.

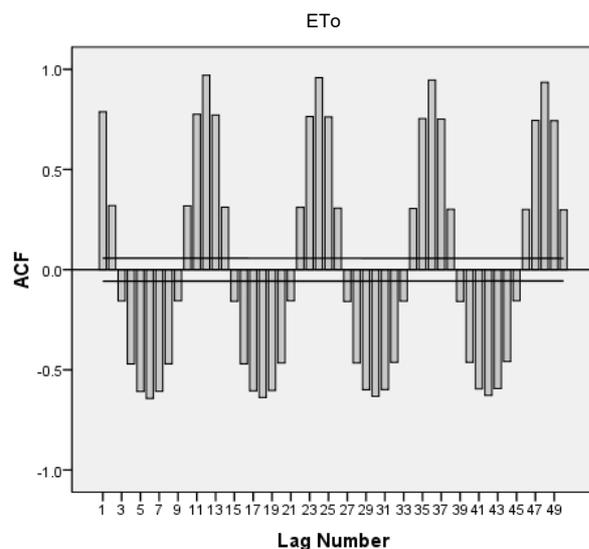


Fig. 1. Auto correlation function (ACF) of Bokaro reference crop evapotranspiration (ET_o) up to 50 lags; source: own study

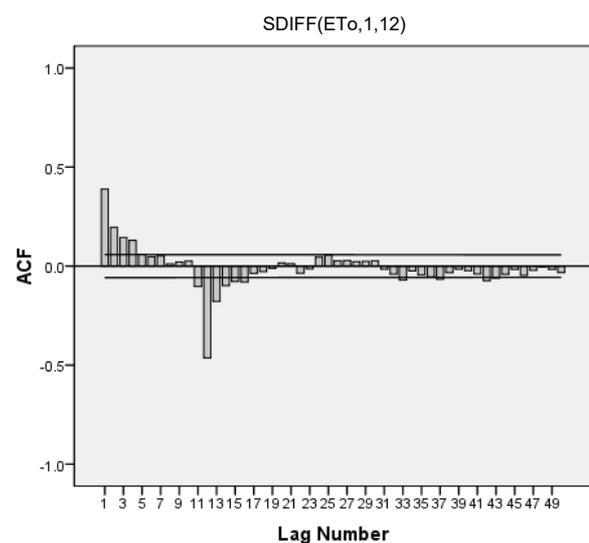


Fig. 3. Auto correlation function (ACF) of Bokaro reference crop evapotranspiration (ET_o) up to 50 lags after seasonal differencing (SDIFF($ET_o,1,12$)); source: own study

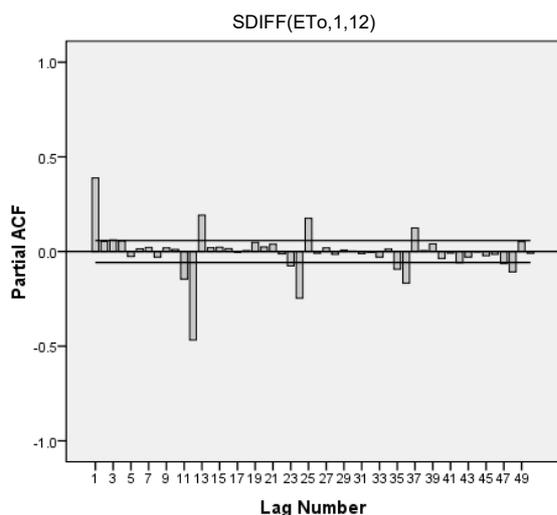


Fig. 4. Partial autocorrelation function (PACF) of Bokaro reference crop evapotranspiration (ETo) up to 50 lags after seasonal differencing (SDIFF(ETo,1,12)); source: own study

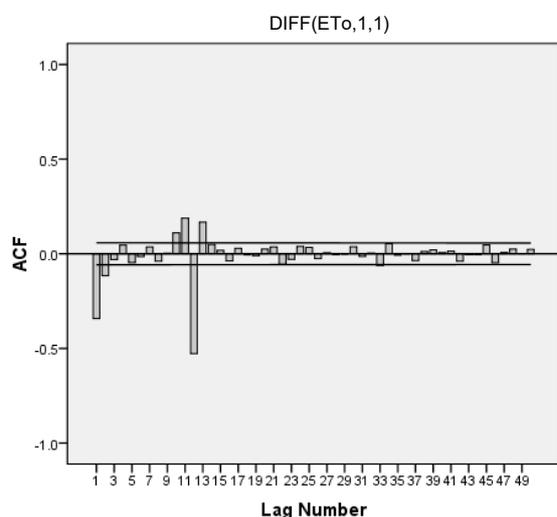


Fig. 5. Autocorrelation function (ACF) of Bokaro reference crop evapotranspiration (ETo) up to 50 lags after seasonal and non-seasonal differencing (DIFF(ETo,1,1)); source: own study

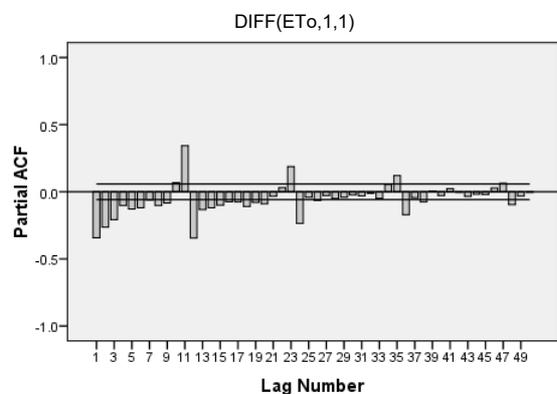


Fig. 6. Partial autocorrelation function (PACF) of Bokaro reference crop evapotranspiration (ETo) up to 50 lags after seasonal and non-seasonal differencing (DIFF(ETo,1,1)); source: own study

RESULT AND DISCUSSION

Sample autocorrelation function and partial autocorrelation function are used for identification of model. This was done after seasonal and non-seasonal differencing of data series. Then parameters (Φ and Θ) of the model are estimated followed by the SE and t values for adequacy and diagnostic of that model. Iterative maximum likelihood method is used for parameter estimation in which optimum value is based on the minimizing of residual sum of the squared between the observed data and estimated one. As definitions of seasonal autoregressive integrated moving average model ARIMA (p, d, q)(P, D, Q), ARIMA (0, 1, 4)(0, 1, 1)₁₂ is identified. This identified model produces residuals ACF and PACF that are not significantly different from zero i.e. all the values of residual were close to zero and at 95% confidence level inside the limit. Figure 7 show the plot of residual ACF and PACF. Table 1 presents the results of parameter values of the model at different lags. In this table ratio of parameter estimated and standard error is greater than 2, so model is adequate, t values greater than 2 also presents adequacy of model. Table 2 presents the statistics of the fitted model, for selected model value of stationary R^2 , R^2 , $RMSE$, $MAPE$, $MaxAPE$, MAE , $MaxAE$, normalized BIC, Q -value are within the satisfactory limit.

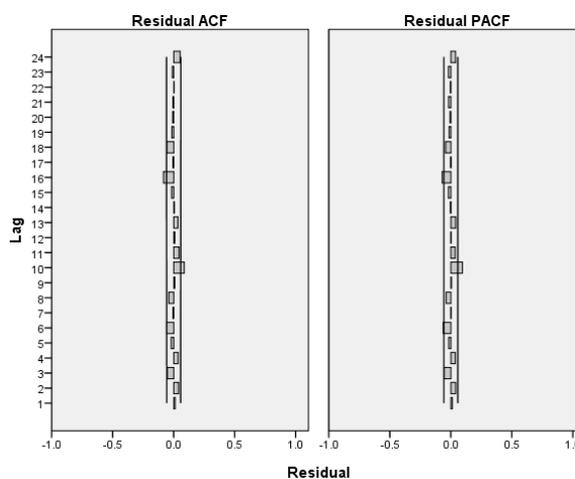


Fig. 7. Residual ACF & PACF of Ranchi ET0; source: own study

Table 1. ARIMA model parameter

Particular	Parameter MA (Θ)	Standard error (SE)	t -test
Moving average (MA) Lag1	0.570	0.028	20.451
Moving average (MA) Lag2	0.193	0.031	6.162
Moving average (MA) Lag4	0.110	0.025	4.470
Moving average (MA) Lag1 seasonal	0.989	0.019	53.292

Source: own study.

Table 2. Statistics of fitted model

Fit statistics	Mean
Stationary R^2	0.618
R^2	0.618
Root mean square error (RMSE)	0.184
Mean absolute percentage error (MAPE)	1.740E14
Maximum absolute percentage error (MaxAPE)	4.779E16
Mean absolute error (MAE)	0.127
Maximum mean absolute error (MaxAE)	1.073
Normalized Bayesian information criterion (BIC)	-0.336
Ljung-Box statistics (Q-value)	38.973

Source: own study.

After selecting the ARIMA model, it is used for forecasting. 24 months ET_o was forecasted by using selected model. Figure 8 shows the scatter plot between observed and forecasted ET_o . The R^2 value 0.9821 presents good correlation between observed and forecasted value. Forecasting can be done for more months but when forecasting months are increases then R^2 values are decreases. In this way model identification, model estimation, diagnostic checking of model and forecasting are the complete processes of the model. Autocorrelation function and partial autocorrelation function play important role in the modelling.

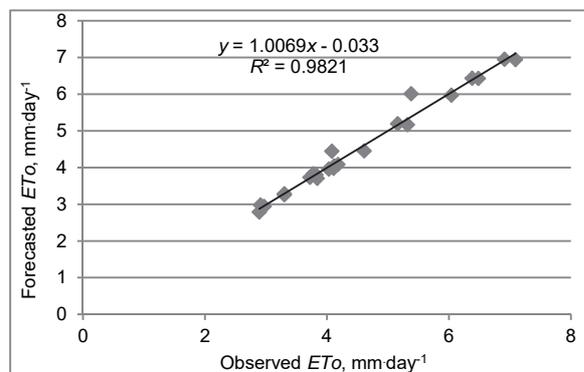


Fig. 8. Scatter plot between observed and forecasted ET_o ; source: own study

CONCLUSION

Time series analysis of evapotranspiration was conducted to help the water system managers and decision makers to establish appropriate strategies to manage and sustain water resources. In time series, by studying the past better decision for future can be made. In this paper reference crop evapotranspiration is forecasted without using weather data. Twelve hundreds month's evapotranspiration data were used for study to ensure satisfactory estimate of monthly values. Seasonal autoregressive integrated moving average modelling has demonstrated good results and performing 24-months forecasting with accuracy lev-

el. Nevertheless forecasting technique presented in this paper allow forecasting for more months but accuracy may be decrease. These finding can be helpful for making local and national water policy, and for irrigation scheduling. This model is also suitable for this type of similar geographic and climatic region. Future research may be addressed to extend for validation data set and validate our finding for other regions. It should explore how to improve forecasting and estimate climate change effects.

Acknowledgements

We are thankful to Birla Institute of Technology, Mesra, Ranchi, India for providing us Research Facilities and grateful to Prof. Gopal Pathak Head of Department, Civil and Environmental Engineering Department, Birla Institute of Technology, Mesra, India for encouraging me continuously for fulfilling the research work.

REFERENCES

- ALHASSOUNL S., SENDILL U., AL-OTHTMAN A.A., NEGM A.M. 1997. Stochastic generation of annual and monthly evaporation in Saudi Arabia. *Canadian Water Resources Journal*. Vol. 22. No. 2 p. 141–154.
- ASADI A., VAHDAT S.F., SARRAF A. 2013. The forecasting of potential evapotranspiration using time series analysis in humid and semi humid regions. *American Journal of Engineering Research*. Vol. 2 p. 296–302.
- BABAZADEH H., SHAMSNLA S.A. 2014. Modelling climate variables using time series analysis in arid and semi-arid regions. *African Journal of Agricultural Research*. Vol. 9 (26) p. 2018–2027.
- ETUK E.H., NATAMBA B. 2015. Modelling monthly Ugandan Shilling/US Dollar exchange rates by seasonal Box-Jenkins techniques. *International Journal of Life Science and Engineering*. Vol. 1. No. 4 p. 165–170.
- GORANTIWAR S.D., PATIL P.D. 2009. Stochastic modelling for crop evapotranspiration for Rahuri Region. *International Journal of Agricultural Engineering*. Vol. 2. No. 1 p. 140–145.
- HAMDI M.R., BDOUR A.N., TARAWNEH A.N. 2008. Developing reference crop evapotranspiration time series simulation model using class a pan: A case study for the Jordan Valley, Jordan. *Journal of Earth and Environmental Sciences*. Vol. 1. No. 1 p. 33–44.
- India Waterportal [dateless]. Met data [online]. [Access 08.04.2016]. Available at <http://indiawaterportal.org/metdata>
- KOTTEGODA N.T. 1980. *Stochastic water resources technology*. London. Mc Millan. ISBN 0470989750 pp. 384.
- MOHAN S., ARUMUGAM N. 1995. Forecasting weekly reference crop evapotranspiration series. *Hydrological Sciences Journal*. Vol. 40. Iss. 6 p. 689–702.
- PARERA A., WESTERN A., NAWARATHNA B., GEORGE B. 2013. Forecasting daily reference evapotranspiration for Shepparton, Victoria, Australia using Numerical Weather Prediction outputs. [20th International Congress on Modelling and Simulation]. [Adelaide, Australia, 1–6 December 2013].
- POPALE P.G., GORANTIWAR S.D. 2014. Stochastic generation and forecasting of weekly rainfall for Rahuri Region. *International Journal of Innovative Research in Science, Engineering and Technology*. Vol. 3. Spec. Iss. 4 p. 185–196.

- PSILOVIKOS A., ELHAG M. 2013. Forecasting of remotely sensed daily evapotranspiration data over Nile Delta Region, Egypt. *Water Resources Management*. Vol. 27 p. 4115–4130.
- SALAS J.D., DELLEUR J.W., YEVJEVIVH V., LANE W.L. 1980. *Applied modeling of hydrologic time series*. Littleton, Colorado. Water Resource Publications. ISBN 0918334 373 pp. 484.
- VALIPOUR M. 2012. Ability of Box-Jenkins models to estimate of reference potential evapotranspiration (A case study: Mehrabad Synoptic Station, Tehran, Iran). *IOSR Journal of Agriculture and Veterinary Science*. Vol. 1. Iss. 5 p. 1–11.
- XU J.Z., PENG S.Z., YANG S.H., LUO Y.F., WANG Y.J. 2012. Predicting daily reference evapotranspiration in a humid region China by the locally calibrated Hargreaves–Samani equation using weather forecast data. *Journal of Agricultural Science and Technology*. Vol. 14 p. 1331–1342.

Ratnesh GAUTAM, Anand K. SINHA

Analiza serii czasowych ewapotranspiracji potencjalnej upraw w dystrykcie Bokaro, Jharkhand, Indie

STRESZCZENIE

Ewapotranspiracja jest jednym z głównych elementów obiegu wody. Dokładniejsze pomiary i możliwość prognozowania ewapotranspiracji mogłyby umożliwić wydajniejsze zarządzanie zasobami wodnymi. Dlatego prezentowane w niniejszej pracy badania skoncentrowane były na modelowaniu i prognozowaniu ewapotranspiracji, ponieważ prognozowanie zapewni więcej informacji do optymalnego zarządzania zasobami wodnymi. Istnieje wiele technik prognozowania ewapotranspiracji, takich jak autoregresja (AR), średnia ruchoma (MA), autoregresyjna średnia ruchoma (ARMA), autoregresyjna zintegrowana średnia ruchoma (ARIMA), metoda Thomasa–Feiringa i inne. Stwierdzono, że spośród nich ARIMA jest bardziej odpowiednia do analizy i prognozowania zdarzeń hydrologicznych. Z tego powodu wykorzystano model ARIMA do prognozowania miesięcznych średnich wartości ewapotranspiracji potencjalnej poprzez analizę stochastyczną. Do analiz i prognozowania użyto serii danych ze 102 lat (1224 miesiące) z dystryktu Bokaro. Na podstawie funkcji autokorelacji (ACF) i cząstkowych autokorelacji (PACF) serii danych wybrano różny porządek modelu ARIMA. Do wyznaczenia parametrów modelu wykorzystano metodę maksymalnego prawdopodobieństwa. Najlepiej dostosowanymi parametrami statystycznymi modelu okazały się ARIMA (0, 1, 4) (0, 1, 1)₁₂.

Słowa kluczowe: ewapotranspiracja, model ARIMA, prognozowanie, serie czasowe